USSN 09/549,848
B2

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 033 405 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
06.09.2000 Bulletin 2000/36

(21) Application number: **00301439.6**

(22) Date of filing: **25.02.2000**

(51) Int Cl.[7]: **C12N 15/29**, C12N 15/82, C07K 14/415, C12Q 1/68, A01H 5/00

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **25.02.1999 US 121825 P**
27.07.1999 US 145918 P
28.07.1999 US 145951 P
02.08.1999 US 146388 P
02.08.1999 US 146389 P
02.08.1999 US 146386 P
03.08.1999 US 147038 P
04.08.1999 US 147302 P
04.08.1999 US 147204 P
More priorities on the following pages

(83) **Declaration under Rule 28(4) EPC (expert solution)**

(71) Applicant: **Ceres Incorporated**
Malibu, CA 90265 (US)

(72) Inventors:
- **Alexandrov, Nickolai**
Thousand Oaks, CA 91320 (US)
- **Brover, Vyacheslav**
Calabasas, CA 91302 (US)
- **Chen, Xianfeng**
Los Angeles, CA 90025 (US)
- **Subramanian, Gopalakrishnan**
Moorpark, CA 93021 (US)
- **Troukhan, Maxim E.**
South Pasadena, CA 91030 (US)
- **Zheng, Liansheng**
Creve Coeur, MO 63141 (US)
- **Dumas, J.**
, (US)

(74) Representative:
Bannerman, David Gardner et al
Withers & Rogers,
Goldings House,
2 Hays Lane
London SE1 2HW (GB)

Remarks:
THE COMPLETE DOCUMENT INCLUDING REFERENCE TABLES AND THE SEQUENCE LISTING IS AVAILABLE ON CD-ROM FROM THE EUROPEAN PATENT OFFICE, VIENNA SUB-OFFICE.

(54) Sequence-determined DNA fragments and corresponding polypeptides encoded thereby

(57) The present invention provides DNA molecules that constitute fragments of the genome of a plant, and polypeptides encoded thereby. The DNA molecules are useful for specifying a gene product in cells, either as a promoter or as a protein coding sequence or as an UTR or as a 3' termination sequence, and are also useful in controlling the behavior of a gene in the chromosome, in controlling the expression of a gene or as tools for genetic mapping, recognizing or isolating identical or related DNA fragments, or identification of a particular individual organism, or for clustering of a group of organisms with a common trait.

[0]Arabidopsis DNA is used in the present experiment, but the procedure is a general one.

EP 1 033 405 A2

[2347] The suspension culture cells are transformed with exogenous DNA as described by Z. Chen et al. *Plant Mol. Bio.* 36:163 (1998). Briefly, 4-days post-subculture cells are incubated with cell wall digestion solution containing 0.4 M sorbitol, 2% driselase, 5mM MES (2-[N-Morpholino] ethanesulfonic acid) pH 5.0 for 5 hours. The digested cells are pelleted gently at 60 xg for 5 min. and washed twice in W5 solution containing 154 mM NaCl, 5 mM KCl, 125 mM $CaCl_2$ and 5mM glucose, pH 6.0. The protoplasts are suspended in MC solution containing 5 mM MES, 20 mM $CaCl_2$, 0.5 M mannitol, pH 5.7 and the protoplast density is adjusted to about 4 x 10^6 protoplasts per ml.

[2348] 15-60 µg of plasmid DNA is mixed with 0.9 ml of protoplasts. The resulting suspension is mixed with 40% polyethylene glycol (MW 8000, PEG 8000), by gentle inversion a few times at room temperature for 5 to 25 min. Protoplast culture medium known in the art is added into the PEG-DNA-protoplast mixture. Protoplasts are incubated in the culture medium for 24 hour to 5 days and cell extracts can be used for assay of transient expression of the introduced gene. Alternatively, transformed cells can be used to produce transgenic callus, which in turn can be used to produce transgenic plants, by methods known in the art. See, for example, Nomura and Komamine, *Pit. Phys.* 79: 988-991 (1985), *Identification and Isolation of Single Cells that Produce Somatic Embryos in Carrot Suspension Cultures.*

[2349] The invention being thus described, it will be apparent to one of ordinary skill in the art that various modifications of the materials and methods for practicing the invention can be made. Such modifications are to be considered within the scope of the invention as defined by the following claims.

[2350] Each of the references from the patent and periodical literature cited herein is hereby expressly incorporated in its entirety by such citation.

Claims

1. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which encodes an amino acid sequence exhibiting at least 40% sequence identity to an amino acid sequence encoded by

   (a) a nucleotide sequence described in REF and/or SEQ Table 1 or 2 or a fragment thereof; or
   (b) a complement of a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof.

2. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which exhibits at least 65% sequence identity to

   (a) a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof; or
   (b) a complement of a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof.

3. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which exhibits at least 65% sequence identity to a gene comprising

   (a) a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof; or
   (b) a complement of a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof.

4. An isolated nucleic acid molecule which is the reverse of the isolated nucleotide sequence according to any one of claims 1-3, such that the reverse nucleotide sequence has a sequence order which is the reverse of the sequence order of said isolated nucleotide sequence according to any one of claims 1-3.

5. An isolated nucleic acid molecule comprising a nucleic acid capable of hybridizing to a nucleic acid having a sequence selected from the group consisting of:

   (a) a nucleotide sequence which is shown in REF and/or SEQ Table 1 or 2; and
   (b) a nucleotide sequence which is complementary to a nucleotide sequence shown in REF and/or SEQ Table 1 or 2;

   under conditions that permit formation of a nucleic acid duplex at a temperature from about 40°C and 48°C below the melting temperature of the nucleic acid duplex.

6. The nucleic acid molecule according to any one of claims 1-5, wherein said nucleic acid comprises an open reading frame.

7. The isolated nucleic acid molecule of any one of claims 1-5, wherein said nucleic acid is capable of functioning as a promoter, a 3' end termination sequence, an untranslated region (UTR), or as a regulatory sequence.

8. The isolated nucleic acid molecule of claim 7, wherein said nucleic acid is a promoter and comprises a sequence selected from the group consisting of a TATA box sequence, a CAAT box sequence, a motif of GCAATCG or any transcriptoin-factor binding sequence, and any combination thereof.

9. The isolated nucleic acid molecule of claim 7, wherein the nucleic acid sequence is a regulatory sequence which is capable of promoting seed-specific expression, embryo-specific expression, ovule-specific expression, tapetum-specific expression or root-specific expression of a sequence or any combination thereof.

10. A vector construct comprising a nucleic acid molecule according to any one of claims 1-9, wherein said nucleic acid molecule is heterologous to any element in said vector construct.

11. A vector construct according to claim 10 comprising:

(a) a first nucleic acid having a regulatory sequence capable of causing transcription and/or translation; and
(b) a second nucleic acid having the sequence of said isolated nucleic acid molecule according to any one of claims 1-4;

wherein said first and second nucleic acids are operably linked and wherein said second nucleic acid is heterologous to any element in said vector construct.

12. The vector construct according to claim 11, wherein said first nucleic acid is native to said second nucleic acid.

13. The vector construct according to claim 11, wherein said first nucleic acid is heterologous to said second nucleic acid.

14. A vector construct according to claim 10 comprising:

(c) a first nucleic acid having having the sequence of said isolated nucleic acid molecule according to claim 7; and
(d) a second nucleic acid;

wherein said first and second nucleic acids are operably linked and wherein said first nucleic acid is heterologous to any element in said vector construct.

15. The vector construct according to claim 14, wherein said first nucleic acid is native to said second nucleic acid.

16. The vector construct according to claim 14, wherein said first nucleic acid is heterologous to said second nucleic acid.

17. A host cell comprising an isolated nucleic acid molecule according to any one of claims 1-4, wherein said nucleic acid molecule is flanked by exogenous sequence.

18. A host cell comprising a vector construct of any one of claims 10-16.

19. An isolated polypeptide comprising an amino acid sequence

(a) exhibiting at least 40% sequence identity of an amino acid sequence encoded by a sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof; and
(b) capable of exhibiting at least one of the biological activities of the polypeptide encoded by said nucleotide seqence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof.

20. The isolated polypeptide of claim 19, wherein said amino acid sequence exhibits at least 75% sequence identity to an amino acid sequence encoded by a sequence shown in SEQ Table 1 or 2 or a fragment thereof.

21. The isolated polypeptide of claim 19, wherein said amino acid sequence exhibits at least 85% sequence identity

to an amino acid sequence encoded by a sequence shown in SEQ Table 1 or 2 or a fragment thereof.

22. The isolated polypeptide of claim 19, wherein said amino acid sequence exhibits at least 90% sequence identity to an amino acid sequence encoded by a sequence shown in SEQ Table 1 or 2 or a fragment thereof.

23. An antibody capable of binding the isolated polypeptide of any one of claims 19-22.

24. A method of introducing an isolated nucleic acid into a host cell comprising:

(a) providing an isolated nucleic acid molecule according to any one of claims 1-4; and
(b) contacting said isolated nucleic with said host cell under conditions that permit insertion of said nucleic acid into said host cell.

25. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to any one of claims 10-16.

26. A method of modulating transcription and/or translation of a nucleic acid in a host cell comprising:

(a) providing the host cell of claim 24 or 25; and
(b) culturing said host cell under conditions that permit transcription or translation.

27. A method for detecting a nucleic acid in a sample which comprises:

(a) providing an isolated nucleic acid molecule according to any one of claims 1-5;
(b) contacting said isolated nucleic acid molecule with a sample under conditions which permit a comparison of the sequence of said isolated nucleic acid molecule with the sequence of DNA in said sample; and
(c) analyzing the result of said comparison.

28. The method according to claim 27, wherein said isolated nucleic acid molecule and said sample are contacted under conditions which permit the formation of a duplex between complementary nucleic acid sequences.

29. A plant or cell of a plant which comprises a nucleic acid molecule according to any one of claims 1-4 which is exogenous to said plant or plant cell.

30. A plant or cell of a plant which comprises a nucleic acid molecule according to any one of claims 1-4, wherein said nucleic acid molecule is heterologous to said plant or said cell of a plant.

31. A plant or cell of a plant which has been transformed with a nucleic acid molecule according to any one of claims 1-4.

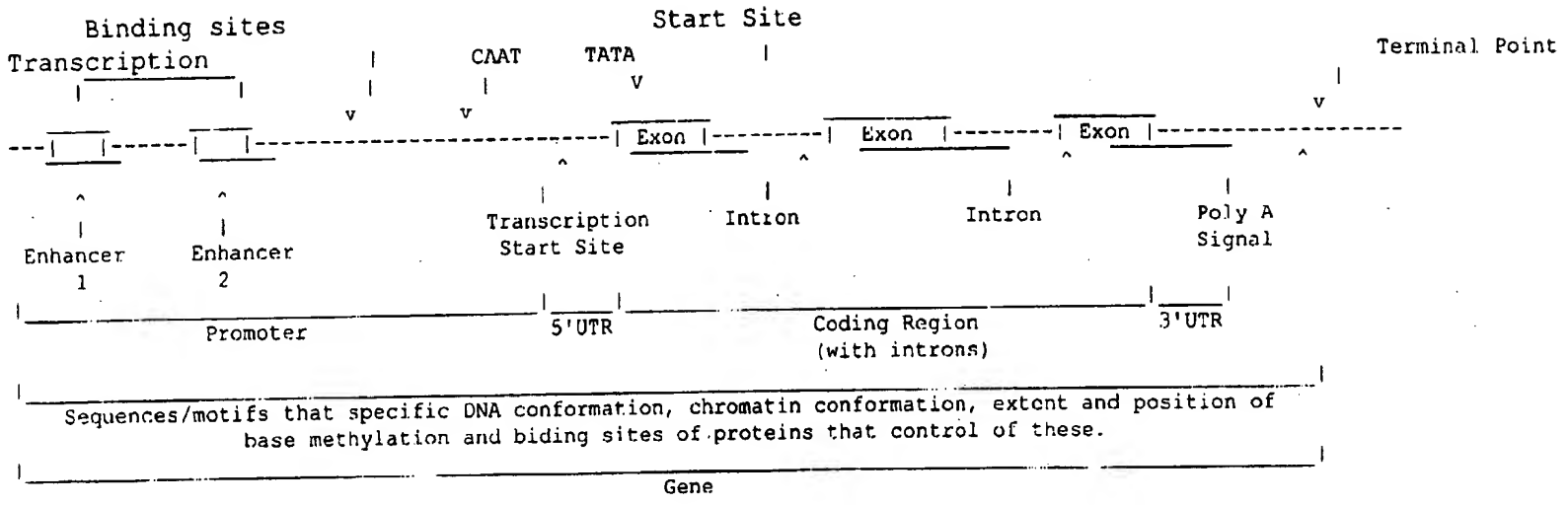
32. A plant of cell of a plant which comprises a vector construct according to any one of claims 10-16.

33. A plant of cell of a plant which has been transformed with a vector construct according to any one of claims 10-16.

34. A plant which has been regenerated from a plant cell according to any one of claims 29-33.

695

SCHEMATIC 1



EP 1 033 405 A2

(Dp) Rel. AA SEQ
- Align. NO 26500
- gi No 5051475
- Desp. : (AL078627) cytochrome c oxidase assembly protein cox10 precursor [Schizosaccharomyces pombe]
- % Idnt. : 37.8
- Align. Len.: 337
- Loc. SEQ ID NO 38170: 96 -> 421 aa.

PolyP SEQ
- Pat. Appln. SEQ ID NO 38171
- Ceres SEQ ID NO 1831973
- Loc. SEQ ID NO 38169: @ 107 nt.

(C) Pred. PP Nom. & Annot.
- Cytochrome c oxidase assembly factor
- Loc. SEQ ID NO 38171: 107 -> 379 aa.

(Dp) Rel. AA SEQ
- Align. NO 26501
- gi No 5051475
- Desp. : (AL078627) cytochrome c oxidase assembly protein cox10 precursor [Schizosaccharomyces pombe]
- % Idnt. : 37.8
- Align. Len.: 337
- Loc. SEQ ID NO 38171: 61 -> 386 aa.

PolyP SEQ
- Pat. Appln. SEQ ID NO 38172
- Ceres SEQ ID NO 1831974
- Loc. SEQ ID NO 38169: @ 428 nt.

(C) Pred. PP Nom. & Annot.
- Cytochrome c oxidase assembly factor
- Loc. SEQ ID NO 38172: 1 -> 272 aa.

(Dp) Rel. AA SEQ
- Align. NO 26502
- gi No 5051475
- Desp. : (AL078627) cytochrome c oxidase assembly protein cox10 precursor [Schizosaccharomyces pombe]
- % Idnt. : 37.8
- Align. Len.: 337
- Loc. SEQ ID NO 38172: 1 -> 279 aa.

Max Len. Seq. :
Pub gDNA:
gi No: 2979540
Gen. seq. in cDNA:

31698	...	31304	OCKHAMG-CDS
31157	...	30635	OCKHAMG-CDS
30355	...	29735	OCKHAMG-CDS

(Ac) cDNA SEQ
- Pat. Appln. SEQ ID NO: 38173
- Ceres SEQ ID NO: 1832050

PolyP SEQ
- Pat. Appln. SEQ ID NO 38174
- Ceres SEQ ID NO 1832051

<210> 38171
<211> 434
<212> PRT
<213> Arabidopsis thaliana
<220>
<223> any n or Xaa = unknown
<220>
<223> Location 1..434 / Ceres Seq. ID 1831973
<400> 38171

```
Met Trp Arg Arg Ser Val Val Tyr Arg Phe Ser Ser Arg Ile Ser Val
1               5                  10                  15
Ser Ser Ser Leu Pro Asn Pro Arg Leu Ile Pro Trp Ser Arg Glu Leu
            20                  25                  30
Cys Ala Val Asn Ser Phe Ser Gln Pro Pro Val Ser Thr Glu Ser Thr
        35                  40                  45
Ala Lys Leu Gly Ile Thr Gly Val Arg Ser Asp Ala Asn Arg Val Phe
    50                  55                  60
Ala Thr Ala Thr Ala Ala Ala Thr Ala Thr Ala Thr Thr Gly Glu Ile
65                  70                  75                  80
Ser Ser Arg Val Ala Ala Leu Ala Gly Leu Gly His His Tyr Ala Arg
                85                  90                  95
Cys Tyr Trp Glu Leu Ser Lys Ala Lys Leu Ser Met Leu Val Val Ala
            100                 105                 110
Thr Ser Gly Thr Gly Tyr Ile Leu Gly Thr Gly Asn Ala Ala Ile Ser
        115                 120                 125
Phe Pro Gly Leu Cys Tyr Thr Cys Ala Gly Thr Met Met Ile Ala Ala
    130                 135                 140
Ser Ala Asn Ser Leu Asn Gln Ile Phe Glu Ile Ser Asn Asp Ser Lys
145                 150                 155                 160
Met Lys Arg Thr Met Leu Arg Pro Leu Pro Ser Gly Arg Ile Ser Val
                165                 170                 175
Pro His Ala Val Ala Trp Ala Thr Ile Ala Gly Ala Ser Gly Ala Cys
            180                 185                 190
Leu Leu Ala Ser Lys Thr Asn Met Leu Ala Ala Gly Leu Ala Ser Ala
        195                 200                 205
Asn Leu Val Leu Tyr Ala Phe Val Tyr Thr Pro Leu Lys Gln Leu His
    210                 215                 220
Pro Ile Asn Thr Trp Val Gly Ala Val Val Gly Ala Ile Pro Pro Leu
225                 230                 235                 240
Leu Gly Trp Ala Ala Ala Ser Gly Gln Ile Ser Tyr Asn Ser Met Ile
                245                 250                 255
Leu Pro Ala Ala Leu Tyr Phe Trp Gln Ile Pro His Phe Met Ala Leu
            260                 265                 270
Ala His Leu Cys Arg Asn Asp Tyr Ala Ala Gly Gly Tyr Lys Met Leu
        275                 280                 285
Ser Leu Phe Asp Pro Ser Gly Lys Arg Ile Ala Ala Val Ala Leu Arg
    290                 295                 300
Asn Cys Phe Tyr Met Ile Pro Leu Gly Phe Ile Ala Tyr Asp Cys Glu
305                 310                 315                 320
Ser Trp Gly Leu Thr Ser Ser Trp Phe Cys Leu Glu Ser Thr Leu Leu
                325                 330                 335
Thr Leu Ala Ile Ala Ala Thr Ala Phe Ser Phe Tyr Arg Asp Arg Thr
            340                 345                 350
Met His Lys Ala Arg Lys Met Phe His Ala Ser Leu Leu Phe Leu Pro
        355                 360                 365
Val Phe Met Ser Gly Leu Leu Leu His Arg Val Ser Asn Asp Asn Gln
    370                 375                 380
Gln Gln Leu Val Glu Glu Ala Gly Leu Thr Asn Ser Val Ser Gly Glu
385                 390                 395                 400
Val Lys Thr Gln Arg Arg Lys Lys Arg Val Ala Gln Pro Pro Val Ala
                405                 410                 415
Tyr Ala Ser Ala Ala Pro Phe Pro Phe Leu Pro Ala Pro Ser Phe Tyr
            420                 425                 430
Ser Pro
```